# EVALUATING ADR IN ADMINISTRATIVE AGENCIES*

Craig McEwen

Bowdoin College

Assistant Administrator Smithe, the dispute resolution specialist of Super Agency (SA), finds the values and goals of alternative dispute resolution very attractive. She also faces an increasing backlog of cases while budget problems limit staff resources for new investigators, attorneys and administrative law judges. Moving cases is not her only concern, however. She and her colleagues worry about public confidence government and administrative agencies and the fairness of procedures for dealing with problems, especially for low and middle income individuals and for small businesses. Attracted to the concept of ADR and knowing that it has been advocated to deal with such varied problems, Administrator Smithe wonders whether ADR would work in her agency, especially in light of her inherent skepticism of the hype for trendy changes.

Administrator Smithe asks her able deputy administrator Sam Green to review the empirical research about ADR and to summarize the results with tentative recommendations. She understands that ADR has been used and studied most heavily in courts, but hopes that those experiences can inform her decision-making within the agency. Within a week, a carefully typed memo appears on her desk:

*This paper is a revised and adapted version of a chapter entitled "Evaluating ADR Programs" in Emercring ADR Issues in State and Federal Courts, edited by Frank E. A. Sander. Chicago: Litigation Section of the American Bar Association, 1991.

To: Assistant Administrator Smithe

From: Sam Green

Date: August 15, 1993

Subject: Empirical Research on ADR

First, the good news. Some excellent empirical research exists. Now the bad news. That research focusses almost exclusively on dispute resolution sponsored by courts, not on dispute resolution in administrative agencies. The court dispute resolution literature emphasizes court-annexed arbitration or mediation and only touches on mini-trials, summary jury trials, or early neutral evaluation. Without this growing body of research, the claims for ADR in courts (and elsewhere) would be based on hope, intuition and anecdote only. The same is true with many

of the criticisms. With the research, we can see both real promise and cautions as we plan our own ADR, but we must do some extrapolation to apply the research to our agency, particularly since Superagency has a duel role, running an ALJ program and acting as a litigant in adjudication proceedings.

The research on court-connected mediation (I won't focus on arbitration because it is likely to be used less in the administrative context) is promising but equivocal. It suggests that under the right conditions mediation can promote settlements and even enhance compliance with court orders. It can be fairly inexpensive for a court to administer and may save parties money. Almost uniformly, usersof mediation appear satisfied with the process and typically feel tha'@ it is fair and appropriate.

At the same time, the effects of mediation on such factors as settlement rates, cost savings to parties or courts, caseload pressures, speed of disposition, kind of outcome, and compliance depend on the circumstances of implementation. For example, when in litigation—or the administrative process—does mediation occur? What kinds of cases go to mediation? Who delivers mediation and administers it? The research also suggests that parties are reluctant to enter mediation voluntarily.

There remains much disagreement about how to define and evaluate the quality of justice or access to justice provided by mediation (and by the courts or administrative agencies). However, we do know that disputants value dignified proceedings in which they feel they have been heard and have participated. Also, we still know relatively little about how to choose which cases belong in which kinds of ADR.

In sum, the research makes clear that many of the claims for mediation are supported by evidence in at least some circumstances, but not all. I recommend that you go forward, but that the agency evaluate its efforts to see how they fare. We can expect that if our mediation or case evaluation is done with competence and dignity users will be satisfied and find it legitimate. However, we cannot be certain about many of the other effects of ADR that are of importance to us.

The research also gives many hints about things we should consider in designing our own programs. Those are too detailed to spell out in this short memo. I also recommend that we send for the Consumer's Guide to-ADR Evaluation Research which I understand has just been printed.

Sam's memo encourages Administrator Smithe to follow her own disposition to try out ADR in her agency, in part because she thinks ADR will help solve some serious problems and in part because she values the ADR programs in their own right. Indeed, she begins to plan a mediation program to deal with internal personnel and equal employment complaints, a mediation program to deal with regulatory enforcement, and a

two-step neutral evaluation/mediation program to deal with contract compliance in agency procurement.

Now Administrator Smithe faces a new set of questions. How should she evaluate these proposed ADR programs through empirical research to see if they are meeting the agency's goals? Which questions could research reasonably address and which not? How can scarce resources be used both for program development and empirical research? Before drawing conclusions, she takes Sam up on his suggestion and orders a copy of the new Consumer's Guide.

## Consumer's Guide To ADR

## Evaluation Research

So you are in the market for some research about your ADR program. Many agency administrators are, but they worry about prices and wonder about which model to buy. Some expect too much of research and others are not sure how it can be used. This issue of our Consumer's Guide helps you sort through the maze of products so that you can choose the evaluation that meets your needs and fits your pocketbook. Our report takes you step by step through some key decisions and problems. Do you need evaluation research at all? What goals do you have for evaluation research? What can empirical research realistically tell you and what can't it tell you? Why do differences in evaluation design matter? Who does evaluation research and at what price? What problems are there in using evaluation research once you have it?

## Do You Need Evaluation Research?

Unfortunately, there is no such thing as a generic ADR evaluation. One good piece of research in one agency in one city and state at one time does not establish the effects of mediation in all places, for all cases and at all times. This is not a fault of the research but rather a consequence of the enormous richness and variability of ADR and of the agencies (or other contexts) that ADR programs reside in. Even ten good evaluation projects about mediation capture only a small range of the variation in its applications. Thus, there are no guarantees that the results from other studies will apply to your program.

You can, however, look for patterns to see how consistent other research findings are—that is to see whether or not they are repeated under widely varying conditions. For example, evidence of high levels of "user satisfaction" comes from dozens of studies of rather different kinds of ADR in varied settings.

Less well established results include, for example, the finding that mediation enhances compliance rates. Four separate research projects in court-related small claims mediation in Maine, Ontario, Oregon and New Jersey show strikingly similar results despite variations in the courts and mediation procedures. However, four sets of findings may not be sufficient to establish with confidence their generalizability. And even if these findings were Applicable to most day-of-trial, small claims mediation in courts where compliance is a common problem, would they apply to mediation of contract compliance issues between a federal agency and a major supplier or to problems surrounding enforcement of environmental regulations where the obligations for compliance often stretch out over years and are technically complex?

The accumulating academic and evaluation research on ADR continues to build understanding of its limits and applications. But you probably cannot wait for the steady but slow growth of the ADR knowledge base, especially since the existing research focuses almost exclusively on court-related ADR programs. If you start your own ADR program in an administrative agency, many of your decisions will be better informed if you gather empirical evidence about the operation and consequences of your own program.

The more information that you gather about your ADR program, the greater the contribution you will make to this growing knowledge base. But more information is not always better when your resources are limited and your central goals have to do with policy choices. Think carefully about the particular uses you have for empirical research.

## What Are Your Evaluation Needs and Goals?

The biggest problem our readers report is that they did not have a clear plan for the use of their evaluation before undertaking it. As a result many evaluations sit on shelves and

collect dust. The evaluation research that significantly affects policy decisions remains the exception rather than the rule. To illustrate the problems read the story below of just such an evaluation reported by an anonymous reader:

***"Wouldn't It Be Interesting to Know? The Story of an Evaluation"***

> The Agency Steering Committee planning and implementing a Pilot ADR program found themselves charged to evaluate it by the outside agency that granted funds for it. Many Committee members viewed the evaluation research as documentation, proof to others of the obvious goodness of the project. For others, it was merely an obligation to fulfill. And for all it helped to satisfy curiosity. What then to study, especially given the wide range of expected or hoped for effects? The resulting research ranged widely across the attitudes of ADR users, the time and costs to the agency and other users, case processing time, settlement rates and speed, the effects in sharpening or narrowing issues, consequences for

continuing relationships among disputants, and the impact on case backlog in the agency. In fact, the guiding question in shaping the research was "wouldn't it be interesting to know whether... 11 Not once did anyone suggest any clear criteria that the program had to meet in order to be judged successful.

The results were almost inevitable—the research documented that people liked the process and that it marginally increased the speed of settlement of ADR cases, but not the overall rate of settlement in the agency. There was some indication that even if the process failed to achieve settlements, it more often helped parties to narrow their conflicts than to escalate the conflict. Cost evidence was unclear. Although the research did not show it, administrators liked the program because it provided them another option for case referral, and each administrator knew of several cases settled,through the process.

How does the committee now evaluate the program? is it to be funded in the next budget cycle? Should it be expanded to other parts of the agency? Should the ADR occur later or earlier'in the process? How does the research relate to these decisions?

Wouldn't it be interesting to know?

There are many possible uses for empirical research about ADR—to meet the requirements of a granting agency; to advance knowledge about dispute resolution; to direct life or death decisions about a program; to persuade skeptics to fund or support ADR; to assist in adjusting and improving the program. Which are your goals?

Foundations and granting agencies often require empirical evaluations to encourage accountability and to increase understanding of ADR. No single study contributes profoundly to that knowledge base, which grows incrementally through the accumulation of scores of research projects. Thus, although each new study is relatively unimportant by itself, it is vitally important to the larger enterprise. The incentives to contribute to that large enterprise may be small, however. Thus, many look to more direct pay-offs from research.

One of these could be viewed as the central purpose of evaluations—life or death Program decisions. These rarely seem to be made on the basis of empirical research, however. Policy-makers seldom set clear enough criteria to permit that to happen—for example, 60% settlement and 80% satisfaction rates or the program is discontinued. Standards of judgment and comparisons are difficult to establish without being completely arbitrary (see below), and, as in the story of the Steering Committee above, programs frequently show up well on some criteria while doing less well on others.

As our readers know, the interests and judgments of important groups such as administrators, advocacy and interest groups, or legislators shape policy choices. ADR may improve the quality of work-life, have anecdotal support, increase scarce resources, be politically popular, embody important public values. Under these circumstances, the

evidence from evaluation research may be used primarily to Justify or defend a program with "scientific data."

In fact, administrative agency ADR programs, like the existence of administrative adjudication, can be viewed as virtuous by definition. ADR provides greater resources for conflict resolution, adds diversity to the system, incorporates values of efficiency and/or consensual and participatory dispute resolution. If these are enough to justify its existence, the question for evaluation might shift from whether or not to do a particular type of ADR to how to do it most effectively.

Empirical research could thus be employed for managerial purposes to help identify problems with an ADR program and permit their correction. Do not assume, however, that routine monitoring through collection and reporting of data about cases or case flow itself constitutes managerial evaluation. Evaluation is a state of mind after all. That state of mind includes a disposition to assess the worth of practices in the light of evidence And to make changes in response to that assessment.

How much help can you expect from empirical research in assessing quality and making decisions about the allocation of resources for ADR, whether they are life or death, public relations, or managerial choices?

## What Evaluation Research Can and Can’t Do

Users should not expect that empirical research will make policy decisions for them. Is mediation better than administrative hearings? Does it work or doesn’t it? Should we continue it or shouldn’t we? Is it better to add a mediation program or create two new administrative law judgeships? Are our mediation sessions too abbreviated or too long? Are our mediators sufficiently trained?

These sorts of policy questions are the ones that motivate evaluation research, but the research itself will not answer them. These choices may profit from evidence, but above all they require complex political and value judgments using multiple and sometimes inconsistent criteria. In other words, evaluation research does not evaluate; decision-makers do. If you embark on evaluation research, be prepared to make evaluative decisions based on the evidence you gather.

You should also not expect that your evaluation research will itself provide all that you need to know about the degree to which your program is delivering high quality dispute resolution. There are at least two reason for that limitation. First, not everyone agrees on what high quality means. That means that there might be disagreement about what to measure, and measurement is one of the central challenges of evaluation research. Some, for example, believe that participant satisfaction is a major indicator of quality of justice while others heartily disagree. These latter folks may insist that disputants do not have the experience to judge the quality of a mediation process; instead they might argue that experts should observe mediation sessions and assess the fairness of procedures.

This brings us to the second limit in assessing program quality. The kinds of data collected through empirical research limit the angles of vision on any program. Thus, questionnaires completed by disputants and observations of mediation sessions tell nothing about the extent and character of mediator training or of the ethical standards for mediators. These latter features might better be evaluated by knowledgeable mediators who could review program procedures, training curricula, and ethical standards (at least on paper). Thus, some dimensions of quality might be ascertained by a "program review" somewhat similar to the accreditation reviews that schools undergo. Such program reviews are not the same as evaluation research.

Quality is such a central issue that a failure to address it could leave your evaluation based largely on cost and efficiency criteria. These latter criteria, by the way, are ones where the evidence is particularly mixed about mediation's effectiveness. It's up to you to help decide how your evaluation should be designed to address these significant quality issues.

Some hints about measuring quality may thus be in order. Settlement rates are the easiest to measure, but they may tell little about quality. Relatively low settlement rates, for example, may suggest that mediation or neutral evaluation is not "working" in producing settlements, and high settlement rates may appear to spell success. At the same time, however, modest rates of non-settlement (20% to 50%) may provide some assurance that the process is not so pressured that parties are being coerced into settlement. Participant satisfaction with outcomes along with their sense of fairness and dignity of proceedings ("procedural justice") may be important but more costly dimensions of quality to measure. To do so will require some form of questionnaire or interviews. The quality or fairness of settlements or outcomes may be difficult to judge in a rigorous scientific fashion, but a small sample of outcomes might be reviewed for adequacy and fairness by a panel of "experts" (although recognize that experts themselves may disagree about how to evaluate outcomes). Similarly, occasional observation of mediation sessions can help provide an independent check on perceptions of participants about the degree to which they participate in the processes, are protected from abuse and inordinate pressure within them, and appear to experience free choice about settlement.

Those who seek evaluation research should expect it to answer most clearly narrow, factual questions. Do mediation participants see the process as fair or not? How long does it take for cases in ADR to reach conclusion? What proportion of cases in ADR settle? Does agency delay decrease or not when ADR is introduced? What the answers to these specific questions mean in broad evaluative terms—does the program work? Is it worth the effort and cost? -- is much harder to say. Also be aware that some apparently simple questions—does ADR save money for the agency? -- are in fact complex and involve difficult comparisons (see insert below).

Be careful if you want to introduce comparisons and confident assertions of causation into the equation. Do mediation cases take less long than cases following the regular administrative track? Are mediated cases more likely to settle than regular administrative track cases? Does the addition of ADR cause the agency backlog to decrease. Does mediation cause the settlement of cases that would otherwise have been resolved through an administrative law hearing?

The ease or difficulty of answering these questions depends largely on the design of your research project. That design has to deal with the compared to what problem—perhaps the central challenge of evaluation research. The degree of confidence you want to have in these comparisons and causal claims affects substantially the costs and complexity of the research.

## The Design of Your Evaluation Research

Comparisons permit judgments of relative effectiveness and can help test assumptions about cause and effect relationships. The challenge of making the right comparisons lies at the center of the design of evaluation research.

Comparisons, and thus standards of judgment, about the results of ADR come from several places. Intuitive standards sometimes make sense, but they are often arbitrary. A settlement rate of 20% may seem low; user satisfaction rates of 45% sound low (but is 65%?). Such intuitive standards are vital management tools but rarely invoked in the evaluation industry.

Other standards come from the accumulated findings of research done in jurisdictions with comparable programs. A small claims mediation settlement rate of 65% in a court program appears to be well within a range of acceptability established by other programs, for example. Relying on such comparisons could simplify evaluation research enormously but may be misleading because of the substantial differences in kinds of cases, in program design, and in agency contexts.

A third source of comparison must come from the evaluation research itself. Are mediation participants more or less likely to perceive their treatment to be fair than those not receiving mediation services in the same agency? Are settlement rates higher or lower in the mediation track than in the regular administrative track? Are costs higher or lower for litigants in mediation or the regular track? Does it cost less to the agency to mediate than to proceed through the regular administrative process? (Caution on cost comparisons: see the special insert on the next page). Making such comparisons meaningful is at the heart of good evaluation research.

To be confident about the causes of what you observe also requires good comparisons. Suppose that your evaluation finds that 80% of the cases in a mediation program settle. Is it possible, however, that mediation was not the cause of settlement but that these cases

would have settled on their own? Might high rates of settlement in a voluntary program result from the settlement-proneness of the parties who choose to participate, not from the processes themselves?

Or suppose that an agency case backlog begins to shrink about the same time that mediation is introduced? Could the shrinkage be caused instead by new case management systems or by changes in complaint rates in the agency? Answers to these questions require that some comparative evidence be gathered about what happens in the absence of mediation.

## Beware Research Bearing Agency Cost Comparisons

Assessments of the costs and cost savings of ADR are among the most difficult and misleading in ADR evaluation. Typically, the comparison desired is between the costs to the agency of processing cases through alternative dispute resolution and the costs of regular agency procedures. one problem of comparison becomes obvious as soon as you ask how much it costs to process a case in the agency. The answer, of course, depends on how the costs are allocated. Do the costs include the total operating cost of the whole agency and some pro-rating of the capital cost of the agency divided by the total number of cases? How are the total costs divvied up among the varying kinds of cases filed in a agency? How do you deal with the fact that a few cases cost an inordinate amount and many cases cost relatively little? Are average case costs very meaningful? How are the costs distributed among different components of the agency?

Typically, the costs of ADR appear to be easier to compute because they occur as discrete additions for a new program to an existing operational budget. But complications arise here too. Is some portion of the administrative cost of the agency part of the cost of ADR, for example, for referrals to ADR? When existing personnel have ADR added on to existing responsibilities, how does one count costs?

The fact that agencies handle increasing or decreasing caseloads with relatively fixed resources means that agency cost comparisons can be very misleading. If an ADR program

costs $150 per case compared to some estimate of regular agency handling of $400 per case does that mean that the agency saved $250 per case by having ADR? Not likely. Did the agency avoid increased costs of $250 per case by adding ADR? Only if agency budgets are directly proportional to case loads. What does the comparison mean even if it can be made?

Cost comparisons also tend to devalue the intangible goals of ADR. For example, might there be value to mediation of personnel issues apart from any (unlikely) savings of agency costs? Is there value in improving the efficiency of settlement of contract disputes even if you do not increase thp likelihood of settlement? I the prospect of improved continuing relationships worth something—assuming that is a result of mediation?

Finally, it is not so clear what the units of cost comparison are. If ADR diminishes the likelihood of costly, but rare, future court conflict but costs a bit more now "per case" to achieve that goal, is it really more costly?

Clearly, cost comparisons must be made with great caution and care. They are more complicated than they initially appear.

To deal effectively with cost measures in an agency context, it may be useful to consider the proposition that "time is money." Careful use of time accounting measures for the work of regular agency personnel in processing cases and for those delivering ADR services may be the key to making meaningful cost comparisons.

Key decisions in the design phase of a program make it easier or harder to collect good comparative data. These decisions can also have a direct effect on the cost of research. Everything about an evaluation is substantially less difficult and more clear and compelling if you design your ADR program as a true experiment.

Although they may face legal, political, or bureaucratic hurdles in their implementation, experiments are quite simple to set up. Take an agency mediation program. Randomly assign half (or one-third or two-thirds or... ) of all eligible cases to mediation and the others to the regular track. Now, comparisons can be made readily between cases in the mediation track (the experimental group) and those not (the control group) on the assumption that the two groups are otherwise equivalent or comparable.

When an experimental design is used, you may be able to gather data from good agency records sufficient to address many issues of pace and settlement and can do so at relatively low cost. Experiments also have the virtue of being both short-lived and repeatable once the program is routinized. You should know that experimental design requires modest administrative cost (to keep special track of the two groups, for example, and to provide different forms of notice to different cases). It can also, on occasion, provide misleading results (see the Substitution Problem below).

If experimental design proves impossible and all cases are eligible for an ADR process, someone must construct a comparison group in a way that convinces skeptics that it is very similar to the ADR group. Cases from another agency location or from another time period are sometimes used, but such techniques can be costly and unreliable. You are much better off using random assignment of cases.

Other basic design issues include the size of samples and the scope of the study. The canons of social research can drive these decisions. Much of the technology of social research has to do with choosing samples, when all cases or disputants cannot be studied. To have "scientifically valid" results, social scientists may want larger and more costly samples than are needed for other purposes. For example, interviews with twenty’ randomly chosen litigants may provide a sufficient check on user perceptions of the process for purposes of a managerial evaluation. Scientific validity may be necessary for

some purposes but not others. Decisions about research design should be made in ways that reflect your purposes and suit your budget.

## Who sells Evaluation Research?

The mystique surrounding evaluation research often clouds user judgments. Publications of lengthy journal articles with complex statistical analyses and reports covering several hundred pages can give the wrong impression about the available evaluation products. The problem is akin to assessing the full range of cars on the market from the middle of a Mercedes-Benz showroom. Contract evaluation research often is required by granting agencies and is typically done by an independent, external (and thus presumably impartial and credible to outsiders) agency such as a leading research firm or occasionally a university; is done on a one-shot basis; costs a great deal of money (in the tens or hundreds of thousands of dollars); takes many months to conclude; is quite comprehensive; and contributes broadly to knowledge in the field.

The importance and prominence of the work of contract evaluation could lead potential buyers of research to overlook do-it-yourself evaluation which is done internally (even with existing staff and resources) and thus may sacrifice credibility; costs much less; produces some results faster; can be on-going rather than one-shot; may be quite narrow in focus; and contributes mainly to local policy or management decisions but can also build knowledge in the field.

Some intermediate level can be achieved by the happy coincidence of your need for research and the interest and availability of faculty and graduate students in local universities or colleges. These academics might help you to carry out a low-cost evaluation or provide advice throughout the course of a do-it-yourself evaluation. Be aware, however, that academic researchers often have their own timetables and agendas (based on a concern with theory, for example), so the research they may want to do could differ some from the work you want done. The choice between contract and or some mix between the two depends on your needs and resources.

## Problems in Using Your Evaluation Research

Our readers report that the use of evaluations once they are completed is not always trouble-free. We summarize their letters in order to warn you of some of the problems. By customizing your evaluation, you may minimize some of these difficulties, but some are simply inherent in the translation of research findings into policy decisions.

Limited comparisons: The results of evaluation research are bounded by the kinds of comparisons available to study. You may compare agency-mediation with the regular administrative judging track. But suppose you want to know whether or not mediation adds more to an agency's capacity to move cases than would expenditures for added administrative judges? You really should compare mediation with the addition of extra administrative judges, but such experiments seem unlikely. Thus, be cautious in your conclusions from research based on limited comparisons.

Rare events: Reliance for comparisons on statistical summaries such as averages disguises the fact that all cases are not alike. In some agencies, cases may not vary enormously in their potential impact on agency resources. In others they do. In these agencies settling one "elephant" case may have far more impact than settling 100 "mouse" cases. Thus, even a statistically insignificant increase in settlement rates could be important if the cases settled included one or two that would have required enormous expenditures of agency resources.

Research may be able to separate out the elephants from the mice, but it seldom has. In fact, the tools of evaluation research work best in assessing patterns among large numbers of cases. They work less well on rare or exceptional events. To deal with this problem you might wish to incorporate the qualitative observations of administrators into the evaluation process. Administrators may be able to help you document "successes" or "failures" in elephant cases.

How much is enough? Even when good comparisons are made, it may be difficult to decide how much difference is enough. Suppose that an ADR innovation has settlement rates of 83% compared to 75% in the regular agency track. Is that a sufficient difference

to justify the expense of the program?

What Did we Really Do? Many evaluations neglect to document what the dispute resolution intervention actually looks like. That is important information to others who may have a different conception of, let us say, mediation than you do. It may also prove important for program administration.

Take the case of an experimental Super Agency ADR program in which lawyer-mediators received three hours of training. Over 50 different lawyers then participated as "mediators." Interviews with some of those lawyers reveal major differences in what they did as neutrals. Some took as long as five or six hours in intensive face-to-face and shuttle mediation. others heard best case presentations and issued an advisory opinion as to case value within an hour. Because evidence suggests these differences had major consequences for settlement, descriptive information on what Actually happened in ADR is essential to program administrators trying to plan future training or to evaluate the utility of "mediation" in this Super Agency program.

To customize your evaluation, you may wish to add or emphasize qualitative components, including observations of dispute resolution processes and in-depth interviews with key actors. Quantitative research is not necessarily more valid or useful than qualitative research.

## Dangers of Relying on A Single Indicator of "Success"

Evaluation can be dangerous to program health when it relies too heavily on a few measures and becomes transformed into performance evaluation. For example, in one state the funding for court mediation programs was directly related to their reported

settlement rates. The higher the rates, the higher the funding. In another, mediators believed that their individual settlement rates were used as the yardstick of their competence. In both instances, the elevation of a single criterion to dominance could encourage other activities that are not monitored such as the use of pressure and coercion to settle in mediation.

The substitution problem. The measurable effects of ADR processes on individual cases may not be the same as their effects on agencies. That is, you might find after a carefully designed experiment that cases in mediation move more quickly than cases without mediation. It would be tempting to conclude that agency backlog or delay will decrease as a result. However, ADR cases often constitute only asmall proportion of cases in a agency and so improving case flow for them may have little is aggregate effect. or alternatively, more time may be devoted to the remaining cases so that they move no more quickly than before.

Innovation effects. -Evaluation research focuses almost entirely on new programs. New ADR programs often are characterized by energy, freshness of perspective, and enthusiastic practitioners. How much of what we see in these programs is a result of these qualities? Is it possible that as programs age they will lose their vigor and some of the characteristics that contribute to positive results? No one really knows, but this concern prompts attention to continued or periodic evaluation through the life of a program.

## The Bottom Line

Evaluation research comes in all shapes and sizes and prices. You probably should have your own because some systematic collection of data can challenge selective perceptions, identify problems for correction, and document successes. Unfortunately, the research never comes with an owner's manual. It is in the translation of research into policy and administrative decisions that unanticipated problems most often arise. Informed by our Consumer's Guide we hope that you can avoid some of these difficulties by being clear about your needs and by knowing the kinds of products and uses for them before you shop.

### *Administrator Smithe Reflects*

Having reviewed Sam's memo and the Consumer's Guide, Administrator Smithe must decide whether and how to evaluate her three ADR programs: the mediation program to deal with internal personnel and equal employment complaints, a mediation program for regulatory enforcement, and a two-step neutral evaluation/mediation program to deal with contract compliance in procurement matters. She has a slim budget for consultants and program development, and virtually no money to carry out extensive research. Her handwritten reflections on her plans follow:

Memo to the File

The enforcement mediation program should reduce our costs, speed cases and not diminish significantly our yield from fines. Most important, it should produce more compliance, less litigation over enforcement and perhaps even fewer repeat offenders. Sam's memo indicated that the research provided some ideas about the conditions under which mediation might speed case movement. I'll ask him to follow up by talking to the Administrative Conference and preparin a detailed memo and a draft proposal for the design of our mediation program. Then we need to call on a couple of consultants who have studied mediation, and ask them to come out here for a day or so to help us think about how we design some research to see if our program has these effects. We should get the research advice before we make final our plan for program operation. Even though we can't afford hiring an outside firm to do the evaluation, we can profit from the advice of some who have done it.

We'll have to rely on do-it-ourselves evaluation of this enforcement mediation effort. I think, though, that we need to design it as an experiment and track cases over time because the comparisons here are important and especially difficult. As we put the program in place, I'll ask the agency administrator to track the cases, reporting settlement rates, time to disposition, fines collected, litigation, and enforcement problems. We will be able to compare the mediation eligible cases (including together both mediated and non-mediated cases) with control group cases on these measures. The mediation group should show differences in its favor on most or all of these measures. In addition, I'll see if we can find a way to examine the expenditure of staff time on cases in each track and compare them. Time expended is a better measure than cost for us.

The personnel/EEO mediation program is a different matter. Although I don't want it to slow cases down, my primary interest is improving the work environment in our agency. What I want is for the agency to make a statement that there should be less contentious ways of resolving workplace disputes, ones that promote effective settlements that meet both employee needs and those of the agency. Because the program will be voluntary and at an early stage of the internal EEO process, the volume of its use will provide a "market evaluation" of its popularity over time, especially of the confidence that employees have in it. Given its newness, it may take a while to catch on, but we can't wait forever. If we don't show a pattern of growth in use over three years, we will definitely have to rethink this program. Ideally, it would also settle cases and cut down on the volume of complaints that require extensive internal investigation and later administrative hearings. But generally, if employees use it and find the process constructive and fair, I think it might be a useful investment of resources.

For evaluation purposes that translates into active monitoring and managerial evaluation. We will not have any control group. We will track voluntary use over time. We will also do periodic exit surveys of participants to assess their views of fairness and satisfaction. At some point, we might ask some knowledgeable outsiders to visit us for a day or two, to review our procedures and program, and to give us some ideas about how we might improve it.

The contract compliance neutral evaluation/mediation rests on the assumption that both our office and outside contractors need a reality check early in a dispute about what might happen if this case went to the Board of Contract Appeals or the Court of Claims. They may also need some help in sharpening issues. Neutral evaluation can help provide that. We will also arrange a later mediation step—after more information is gathered, if necessary—if the parties think it might be fruitful. Some of these cases will be relatively small ones, but some will have huge stakes. We should examine whether or not the size of the contract makes a difference in outcome. Ultimately, we hope that we can reduce our litigation costs and settle this cases more quickly using and efficiently using these two steps.

I think we had better set this one up experimentally too. We can randomly assign cases to mediation or the regular process and then compare the two groups on dimensions such as time to settlement, nature of settlement, continuation/discontinuation of contractual relationship, likelihood of formal proceedings. We hope that the mediation group will be more favorable on most of these dimensions. of course, we also need to see how frequently parties use mediation when it is available to them and try to understand why or why not it is employed.